

**COLINEARITE ENTRE VARIABILITE INDIVIDUELLE
DES LONGUEURS AUX AGES, CROISSANCE MOYENNE
ET STRUCTURES DEMOGRAPHIQUES
DANS L'ANALYSE DES FREQUENCES DE TAILLE**

M. BOER* et A. LAUREC*

Résumé :

Cet article prolonge les travaux de SPARRE (1985), qui généralisaient eux mêmes les articles de SCHNUTE et FOURNIER (1980), FOURNIER et BREEN (1983) au traitement d'un ensemble d'histogrammes. L'originalité du présent document, outre quelques discussions sur le traitement des contraintes, correspond d'une part à des développements analytiques des dérivées premières et secondes de la fonction objectif, d'autre part à l'analyse en composantes principales de l'inverse de son Hessien. Cette analyse permet d'étudier la structure des problèmes de colinéarité. Le traitement d'un exemple emprunté à SPARRE (1985) suggère à cet égard que ces problèmes pourraient être très importants, et limiter sévèrement le nombre de paramètres que l'on peut estimer à partir de seuls histogrammes de distributions de longueur.

Summary :

SPARRE (1985) has generalized the technique initiated by SCHNUTE et FOURNIER (1980), FOURNIER et BREEN (1983) in order to deal with several length frequency distributions, instead of a single histogram. This paper develops the same approach. Beyond some considerations about the possible constraints, the new points

* Laboratoire ERHAL
Centre IFREMER de Nantes
BP 1049 - 44037 Nantes Cedex - FRANCE

correspond on one hand to analytical derivations of the first and second derivatives of the objective function, on the other hand to the analysis of colinearities. It is suggested for the second question to perform a principal component analysis of the invert of the Hessian matrix of the objective fonction. An example, borrowed from SPARRE (1985) is treated, which suggests that colinearity problems could be extremely important, reducing severely the number of parameters one can estimate from length distributions.

1. Introduction

L'analyse des distributions de longueurs pour en déduire une estimation des paramètres de croissance et des structures démographiques bénéficie en halieutique d'une longue tradition historique. Les études se sont développées d'une part selon une voie empirique, dont les derniers développements peuvent être liés aux articles de PAULY et GASCHUTZ (1982), PAULY (1985), d'autre part selon une voie statistiquement plus classique, où le document de base reste celui de HASSELBLAD (1966). Cette seconde approche est celle dont l'analyse est poursuivie ici, selon une philosophie inspirée de SCHNUTE et FOURNIER (1980). Ces auteurs ont en effet tenté de réduire le nombre de paramètres à estimer en astreignant la croissance linéaire à suivre une courbe de VON BERTALANFFY d'une part, en introduisant d'autre part des contraintes sur les relations entre âge et dispersion des longueurs individuelles, éventuellement sur la structure démographiques. Tant dans cet article, que dans ses prolongements (FOURNIER et BREEN, 1983; BREEN et FOURNIER, 1984) un seul histogramme est considéré. SPARRE (1985) a quant à lui proposé une généralisation à plusieurs histogrammes. Cette même approche sera suivie ici.

Les paragraphes 2 à 6 correspondent essentiellement à une discussion de la fonction objectif, du traitement des contraintes, puis à une explicitation de ses dérivées premières et secondes. Ces dérivations permettent une accélération des maximisations, et conduisent pour le passage à l'ordre 2 au Hessien. Ce Hessien donnera accès, au prix d'hypothèses précises, à la matrice de variance-covariance, et de façon plus générale à une analyse des problèmes de colinearité. Compte tenu du nombre élevé de paramètres, l'analyse en composantes principales paraît un outil précieux pour analyser les colinéarités. C'est ce qui est suggéré dans le paragraphe 6, avant qu'une illustration soit offerte par le paragraphe 7. L'ensemble des calculs a été rendu possible par

un programme FORTRAN, faisant un large usage de la bibliothèque NAG.

2. Généralités/Notations

Soit un ensemble de H histogrammes de fréquences de longueurs, indicés par h ($h=1,2, \dots, H$), chaque histogramme correspondant à une date précise. Le pas est supposé constant, commun aux histogrammes, et égal à d . Les âges dans chaque histogramme sont indicés par j ($j=1,2, \dots, G_h$), G_h étant le nombre de modes ou groupes d'âges pour l'histogramme h . On supposera que dans chaque groupe les longueurs suivent une loi normale de moyenne μ_{hj} et d'écart-type σ_{hj} . p_{hj} est la proportion du groupe d'âge j dans l'histogramme h . Pour chaque histogramme, l'indice i est lié à la classe de longueur ($i=1,2, \dots, N$), N étant le nombre de classes de tailles. f_{hi} est l'effectif de la $i^{\text{ème}}$ classe de taille dans l'histogramme h .

Le problème posé est celui de l'estimation des $\{\mu_{hj}\}$, $\{\sigma_{hj}\}$ et $\{p_{hj}\}$ pour $h=1,2, \dots, H$; $j=1,2, \dots, G_h$ en respectant des contraintes telles que celles définies dans le paragraphe 3 ci-après.

3. Les Contraintes

La première contrainte habituellement utilisée porte sur les paramètres $\{p_{hj}\}$. D'autres contraintes portent sur les paramètres $\{\mu_{hj}\}$ et $\{\sigma_{hj}\}$. Deux approches seront utilisées dans cette étude, correspondant à des paramétrages différents.

3.1. Somme des fréquences relatives

De façon évidente:

$$\sum_{j=1}^{G_h} p_{hj} = 1 \quad ; \quad h=1,2, \dots, H; \quad j=1,2, \dots, G_h \quad (1)$$

Ces contraintes peuvent être éliminées par une réduction du nombre de variables, l'un quelconque, P_{h_0} , des P_{h_j} étant déduit des autres par:

$$P_{h_0} = 1 - \sum_{j \neq 0} P_{h_j}$$

3.2. Paramètres L, K, t_0, a, b et $\{P_{h_j}\}$

La première contrainte correspond au souci de respecter pour le déplacement des modes, une équation de croissance du type VON BERTALANFFY. Ainsi:

$$\mu_{h_j} = L - [1 - \exp\{-K(s_{h_1} + j - 1 - t_0)\}] \quad (2)$$

avec les notations usuelles. s_{h_1} étant l'âge du premier groupe d'âge pour l'histogramme h .

La deuxième contrainte porte sur les écarts-types et est reprise de SCHNUTE et FOURNIER (1980), BREEN et FOURNIER (1984), en la modifiant pour couvrir le cas de plusieurs histogrammes. Ainsi:

$$\sigma_{h_j} = a + b \sqrt{(s_{h_j} - s_{11}) / (G_h - s_{11})} \quad (3)$$

a étant l'écart-type pour le premier groupe d'âge, et b étant la différence entre l'écart-type du premier groupe et celui du dernier. s_{h_j} est l'âge précis pour le groupe j et l'histogramme h , en intégrant la date pour chaque histogramme.

La dernière contrainte porte sur la mortalité totale, supposée constante et égale à Z au delà d'un âge dit de plein recrutement. Elle sera écrite:

$$\sum_h \sum_{j \geq t_r} [Z_{h_j} - Z]^2 = 0 \quad (4)$$

où $Z_{h_j} = \log [P_{h_j} / P_{h_{(j-1)}}]$ pour $j \geq t_r$, si t_r est l'indice du premier groupe d'âge pleinement recruté. Une autre forme peut être envisagée, adaptée de FOURNIER et BREEN (1983), et définie comme suit:

$$\sum_h \sum_{j \geq t_r} [\log P_{h_j} - \log c_h + jZ]^2 = 0 \quad (5)$$

Ces contraintes peuvent être éliminées si l'on opère une réduction ~~et une modification~~ du nombre de variables. Ainsi à partir de l'équation (5) on voit qu'il est possible de déduire les p_{hj} des paramètres c_h et Z , en écrivant $p_{hj} = c_h e^{-jZ}$. Les paramètres p_{hj} , pour $j > t_r$ disparaissent ainsi, et sont remplacés par Z d'une part, et les c_h d'autre part.

3.3. Reparamétrage en L_-, L_+, k, a, b et $\{p_{hj}\}$

Le paramétrage ainsi évoqué ne diffère du précédent que pour ce qui touche aux $\{\mu_{hj}\}$, et conduit à l'utilisation des paramètres $\{L_-, L_+, k\}$ au lieu de $\{L_-, K, t_o\}$ (SCHNUTE et FOURNIER, 1980; BREEN et FOURNIER, 1984). Soit $v_{hj} = 1 - k^{S_{hj}-1}$, on posera alors:

$$\mu_{hj} = L_- + (L_+ - L_-) v_{hj} / v_{hgh} \quad (6)$$

On utilise les relations suivantes pour revenir aux paramètres habituellement utilisés:

$$L_- = - \sum_{h=1}^H \frac{L_+ - L_- k^{S_{hgh} - S_{h1}}}{H^n v_{hgh}}$$

$$K = - \log k$$

et

$$t_o = - \sum_{h=1}^H \frac{S_{h1}}{H^n} - \frac{1}{\log k} \log \left(\frac{L_+ - L_-}{L_+ - L_- k^{S_{hgh} - 1}} \right)$$

4. Fonction Objectif

La fonction objectif à maximiser s'écrit sous la forme:

$$g(L_-, K, t_o, a, b, \{p_{hj}\}) = \sum_{h=1}^H \sum_{i=1}^N f_{hi} \log [Q_{hi}] \quad (7)$$

en reprenant la procédure de HASSELBLAD (1966). Elle définit ~~E~~ une approximation de la log-vraisemblance, si les histogrammes analysés correspondent à des échantillonnages aléatoires simples, mutuellement indépendants.

Il a été supposé que pour chaque histogramme, dans chaque groupe d'âge j , les distributions de longueur suivent une loi normale avec pour moyenne μ_{hj} et pour écart-type σ_{hj} . f_{h1} est l'effectif observé dans la classe de taille $[x_{h1}-\frac{1}{2}d; x_{h1}+\frac{1}{2}d]$. L'effectif théorique correspondant est:

$$Q_{h1} = \sum_{j=1}^{G_h} P_{hj} q_{h1j} \quad (8)$$

avec en première approximation:

$$q_{h1j} = \frac{d}{\sigma_{hj}\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left[\frac{x_{h1} - \mu_{hj}}{\sigma_{hj}} \right]^2 \right] \quad (9)$$

où μ_{hj} et σ_{hj} sont définis par les équations (2) et (3), ou les équations (6) et (3), selon le paramétrage utilisé.

On supposera encore que les groupes d'âge~~s~~ correspondent à des âges régulièrement espacés de une unité de temps (en pratique un an).

5. Résolution Numérique

5.1. Premier Paramétrage

5.1.1. Dérivation Préliminaire

Pour expliciter plus facilement les dérivées premières et secondes, quelques notations s'avéreront utiles. On posera ainsi:

$$u = u_{h1j} = (x_{h1} - \mu_{hj}) / \sigma_{hj} = (x - \mu) / \sigma \quad (10)$$

$$t = t_{hj} = s_{h1} + j - 1 - t_0 \quad (11)$$

$$w = w_{hj} = \sqrt{(s_{hj} - 1) / (s_{h0} - 1)} \quad (12)$$

$$Q = Q_{h1} \quad (13)$$

$$q = q_{h1j} \quad (14)$$

et

$$s = s_{hj} \quad (15)$$

Avec les symboles ci-dessus, les indices peuvent être omis pour alléger les formules. Les indices sont alors implicitement ceux donnés dans les formules (10) à (15).

Soit θ l'ensemble des paramètres à estimer:

$$\theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\} = \{L, K, t_0, a, b\} \quad (16)$$

alors

$$\begin{aligned} \beta_m^I &= \delta\mu/\delta\theta_m = 1 - e^{-Kt} ; m=1 & (17) \\ &= tL e^{-Kt} ; m=2 \\ &= -KL e^{-Kt} ; m=3 \\ &= \delta\sigma/\delta\theta_m = 1 ; m=4 \\ &= \omega ; m=5 \end{aligned}$$

et

$$\begin{aligned} \beta_{mn}^{II} &= \delta^2\mu/\delta\theta_m\delta\theta_n ; m=1,2,3; n=1,2,3,4,5; m \leq n & (18) \\ &= \delta^2\sigma/\delta\theta_m\delta\theta_n ; m=4,5; n=4,5; m \leq n \end{aligned}$$

Avec une notation matricielle, on écrira encore:

$$\beta_{mn}^{II} = \begin{pmatrix} 0 & & & & \\ te^{-Kt} & -t^2L e^{-Kt} & & & \\ -Ke^{-Kt} & -L(1-Kt)e^{-Kt} & -K^2L e^{-Kt} & & \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(le triangle supérieur se complétant par symétrie).

Avec la symbolique décrite par les équations (10) à (15) on aboutit à:

$$\delta u/\delta\mu = -1/\sigma \quad \text{et} \quad \delta u/\delta\sigma = -u/\sigma \quad (19)$$

par ailleurs:

$$\delta q/\delta\mu = qu/\sigma \quad \text{et} \quad \delta q/\delta\sigma = q(u^2-1)/\sigma \quad (20)$$

Enfin:

$$Q_m^I = \delta Q/\delta\theta_m ; m=1, \dots, 5 \quad (21)$$

$$Q_{mn}^{II} = \delta^2 Q/\delta\theta_m\delta\theta_n ; m=1, \dots, 5; n=1, \dots, 5; m \leq n \quad (22)$$

et

$$\delta Q/\delta p = q \quad (23)$$

5.1.2. Dérivées Premières

De la définition de g, on déduit immédiatement:

$$\delta g / \delta \theta_m = \sum_n \sum_i f Q_m^i / Q \quad (24)$$

où

$$\begin{aligned} Q_m^i &= \sum_j p q u \beta_m^i / \sigma & m=1,2,3 \\ &= \sum_j p q (u^2-1) \beta_m^i / \sigma & m=4,5 \end{aligned} \quad (25)$$

et

$$\delta g / \delta p = \sum_i f q / Q \quad (26)$$

Si par ailleurs on prend en compte la contrainte $\sum_j p_{n_j} = 1$, pour éliminer $p_{n_j^0}$, calculé par $p_{n_j^0} = 1 - \sum_{j \neq j^0} p_{n_j}$, les dérivées dg/dp_{n_j} se déduisent des dérivées formelles données par l'équation (26) à l'aide de la relation $dg/dp_{n_j} = \delta g / \delta p_{n_j} - \delta g / \delta p_{n_j^0}$. Lorsque l'on traite le problème de la décroissance exponentielle des effectifs, par ~~réduction de~~ changement de variables, on applique de même les formules usuelles sur la dérivation des fonctions de fonctions.

5.1.3. Dérivées Secondes

$$\delta^2 g / \delta \theta_m \delta \theta_n = \sum_n \sum_i f [Q Q_{mn}^i - Q_m^i Q_n^i] / Q^2 \quad (27)$$

avec $m=1, \dots, 5$; $n=1, \dots, 5$; $m \leq n$, où

$$\begin{aligned} Q_{mn}^i &= \sum_j p q [\beta_{mn}^i]^2 [u^2-1] / \sigma^2 & ; m=1; n=1 \\ &= \sum_j p q \beta_n^i [\beta_m^i (u^2-1) + u \sigma / L_{-}] / \sigma^2 & ; m=1; n=2,3 \\ &= \sum_j p q u \beta_m^i \beta_n^i [u^2-3] / \sigma^2 & ; m=1,2,3; n=4,5 \\ &= \sum_j p q \beta_m^i [\beta_n^i (u^2-1) - u \sigma] / \sigma^2 & ; m=2; n=2 \\ &= \sum_j p q \beta_m^i [\beta_n^i (u^2-1) + u K \sigma] / \sigma^2 & ; m=3; n=3 \end{aligned} \quad (28)$$

$$= \sum_j p q \beta_m^I \beta_n^I [(u^2-1)(u^2-2)-2u^2]/\sigma^2; m=4,5; n=4,5; m \leq n$$

$$= \sum_j p q [\beta_m^I \beta_n^I (u^2-1) + u \sigma (K \beta_m^I + \beta_n^I / K)] / \sigma^2; m=2; n=3$$

et

$$\delta^2 g / \delta \theta_m \delta p_{h_j} = \sum_1 f q [Q u \beta_m^I / \sigma - Q_m^I] / Q^2; m=1,2,3 \quad (29)$$

$$= \sum_1 f q [Q (u^2-1) \beta_m^I / \sigma - Q^I h_m] / Q^2; m=4,5$$

$$\delta^2 g / \delta p_{h_m} \delta p_{h_n} = \sum -f q_{h_m} q_{h_n} / Q^2; m > n \quad (30)$$

Si comme précédemment on élimine un des p_{h_j} , des corrections des dérivées formelles seront nécessaires. Elles se feront par les relations:

$$d^2 g / d \theta_m d p_{h_j} = \delta^2 g / \delta \theta_m \delta p_{h_j} - \delta^2 g / \delta \theta_m \delta p_{h_j^0}$$

$$d^2 g / d p_{h_m} d p_{h_n} = \delta^2 g / \delta p_{h_m} \delta p_{h_n} - \delta^2 g / \delta p_{h_m} \delta p_{h_j^0} - \delta^2 g / \delta p_{h_j^0} \delta p_{h_n} + \delta^2 g / \delta^2 p_{h_j^0}$$

5.2. Deuxième paramétrage

5.2.1. Dérivation Préliminaire

On conservera les notations du paragraphe 5.1.1., à l'exception des paramètres de croissance θ . On posera ainsi:

$$\theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\} = \{L_-, L_+, k, a, b\} \quad (31)$$

β_m^I et β_{mn}^II s'écriront en notation matricielle comme:

$$\beta_m^II = \begin{pmatrix} 1 - v_j / v_G \\ v_j / v_G \\ (L_+ - L_-)(v_j v_G - v_j v_G^I) / v_G^2 \\ 1 \\ \omega \end{pmatrix} \quad (32)$$

où $v_{h_j}^I = \delta v_{h_j} / \delta k$, et

$$\beta_{mn}^II = \begin{pmatrix} 0 & & & & \\ 0 & & & & \\ (v_j v_G - v_j v_G^I) / v_G^2 & (v_j v_G^I - v_j v_G) / v_G^2 & (L_+ - L_-) \delta \beta_3^I / \delta k & & \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

où

$$\delta \beta_{\bar{z}} / \delta k = [(v_{j'}^H v_{\sigma} + v_j v_{\sigma}^H) v_{\sigma} - 2(v_j v_{\sigma} - v_j v_{\sigma}^H)] / v_{\sigma}^2$$

et

$$v_{Hj}^H = \delta^2 v_{Hj} / \delta k^2.$$

5.2.2. Dérivées Premières et Secondes

Les équations (24), (27) et (29) seront adaptées en utilisant les équations (32) et (33) au lieu de (16) et (17).

6. Etude des Colinéarités et ACP

Si les histogrammes analysés correspondent à des échantillonnages aléatoires simples, mutuellement indépendants, la fonction objectif correspond à une vraisemblance logarithmique. Dans ce contexte l'inverse du Hessien de la fonction objectif fournit une estimation de la matrice de variance-covariance de l'estimation des différents paramètres. Pour peu que le nombre de paramètres soit élevé, il devient difficile d'apprécier les faits majeurs recélés par cette matrice. Le contexte ainsi décrit relève des techniques d'analyse des données, qui permettent d'offrir des résumés "optimaux" d'un réseau complexe de relations entre variables, tel que traduit par une matrice de variance-covariance, ou de corrélations. Il semble préférable de se ramener à des corrélations, pour éviter les problèmes d'échelle entre les différentes variables. De façon générale les inverses du Hessien seront standardisées pour ramener les termes diagonaux à l'unité.

Si l'on considère ainsi une matrice C de "corrélations" entre variables, l'ACP (Analyse en Composantes Principales) conduit à extraire des projections dans les premiers axes, associés aux premiers vecteurs propres de C, et mesure la pertinence des résumés ainsi créés par les pourcentages de variance expliqués.

La référence faite aux ACP n'implique aucune ambition factorialisante, en termes de facteurs communs et spécifiques. Elle correspond au seul souci d'effectuer des projections sur un sous-espace de dimension réduite, qui respecte "au mieux" la matrice de corrélation donnée. Dans cet esprit la justification la plus

simple de la pertinence de l'ACP peut être trouvée chez BRILLINGER (1975).

Soit une variable X , de dimension N , de matrice de variance-covariance C . L'ACP, par extraction des R premiers vecteurs propres, permet de calculer la projection $D E X$, où D est une matrice ($N \times R$), E une matrice de dimension ($R \times N$), minimisant simultanément toutes les valeurs propres, et donc la trace, de la matrice de variance-covariance de $X - D E X$. Cela aboutit donc à une minimisation globale des résidus $X - D E X$, écarts entre X et la projection $D E X$ sur le sous-espace de dimension R .

L'extraction des vecteurs et valeurs propres a donc été conduite. La saturation de la variable \bullet_j par rapport au facteur k , est ainsi $v_{jk} = u_{jk} \sqrt{\lambda_k}$, si u_{jk} est la $j^{\text{ème}}$ coordonnée du $k^{\text{ème}}$ vecteur propre de C , λ_k étant la valeur propre associée.

Lorsque les hypothèses d'échantillonnage aléatoire simple ne sont pas satisfaites, ce qui hélas sera souvent le cas en pratique, l'inverse du Hessien ne fournit pas directement la matrice de variance-covariance. Il permet néanmoins d'étudier la structure des problèmes de colinéarité. C'est dans cet esprit que nous proposons d'utiliser l'ACP de l'inverse du Hessien.

7. Illustration

7.1. Les données/les traitements

Elles correspondent à deux histogrammes empruntés à SPARRE (1985), qui sont reportés dans le Tableau 1. Ces données ont été générées par simulation selon des hypothèses sous-jacentes, qui sont difficilement compatibles avec l'hypothèse d'une décroissance exponentielle des effectifs au delà d'un âge quelconque. La contrainte correspondante n'a donc pas été incorporée.

Les résultats de l'ajustement apparaissent dans le Tableau 2, et ce pour les deux paramétrages proposés pour la croissance.

Les Hessiens ont été systématiquement calculés par la voie analytique, et vérifiés par les approximations liés aux incréments finis. Il a été choisi de ramener les inconnues sur les structures démographiques aux quatre premiers groupes d'âges. Les effectifs du cinquième se déduisent des premiers par complètement à 100%. Les inconnues se ramènent donc à treize; trois paramètres

de croissance linéaire, deux paramètres de dispersion, et huit (ou deux fois quatre) paramètres de structure démographique.

Tableau 1. Les fréquences observées

Inter valle	Date		Inter valle	Date	
	15-03-1975	20-07-1975		15-03-1975	20-07-1975
8- 9	1	0	32-33	111	53
9-10	11	5	33-34	86	55
10-11	54	21	34-35	96	77
11-12	138	65	35-36	68	69
12-13	156	91	36-37	68	77
13-14	84	52	37-38	63	61
14-15	23	19	38-39	70	69
15-16	3	5	39-40	60	51
16-17	0	8	40-41	55	45
17-18	0	20	41-42	65	38
18-19	0	39	42-43	60	24
19-20	1	77	43-44	51	27
20-21	3	85	44-45	42	18
21-22	5	90	45-46	33	13
22-23	11	69	46-47	29	9
23-24	20	61	47-48	19	8
24-25	32	32	48-49	15	6
25-26	54	17	49-50	10	3
26-27	64	8	50-51	5	2
27-28	68	10	51-52	4	1
28-29	91	18	52-53	2	1
29-30	88	26	53-54	1	0
30-31	99	33	54-55	1	0
31-32	95	43			
			Total	2115	1601

L'analyse a été faite soit sur le Hessien complet, soit en ne prenant en compte qu'une partie des variables. Dans le second cas on opère la minimisation conditionnelle de la fonction objectif, et les variances-covariances sont elles-même conditionnelles.

Tableau 2. Resultats de l'ajustements
avec le parametrage $\theta = (L_-, K, t_0, a, b, (p_{hj}))$ et $\theta = (L_-, L_+, k, a, b, (p_{hj}))$
($h=1,2; j=1, \dots, 5$)

COHORTE	GROUPE (j)	$\theta = (L_-, K, t_0, a, b, (p_{hj}))$ $L_- = 50.0000; K = .3503; t_0 = .4306;$ $a = 2.3449; b = 2.5003$				$\theta = (L_-, L_+, k, a, b, (p_{hj}))$ $L_- = 50.6236; K = .3425; t_0 = .4283;$ $a = 2.2878; b = 2.4266$			
		μ_{1j}	σ_{1j}	p_{1j}	e_{1j}	μ_{1j}	σ_{1j}	p_{1j}	e_{1j}
1975	1	11.8480	2.8941	.2500	1.2027	11.8400	2.8208	.2500	1.2027
1974	2	23.1216	3.6825	.1200	2.2027	23.2943	3.5860	.1200	2.2027
1973	3	31.0640	4.1551	.3281	3.2027	31.4269	4.0446	.3521	3.2027
1972	4	36.6594	4.5276	.0400	4.2027	37.2010	4.4062	.0400	4.2027
1971	5	40.6014	4.8452	.2619	5.2027	41.3006	4.7145	.2379	5.2027

COHORTE	GROUPE	μ_{2j}	σ_{2j}	p_{2j}	e_{2j}	μ_{2j}	σ_{2j}	p_{2j}	e_{2j}
1975	1	16.2254	3.2147	.3254	1.5507	16.0318	3.1320	.3133	1.5507
1974	2	26.2055	3.8045	.2300	2.5507	25.8563	3.7044	.2300	2.5507
1973	3	33.2366	4.2168	.1100	3.5507	32.8317	4.1046	.1100	3.5507
1972	4	38.1901	4.5535	.2946	4.5507	37.7843	4.4313	.3067	4.5507
1971	5	41.6798	4.8452	.0400	5.5507	41.3006	4.7145	.0400	5.5507

7.2. Résultats de l'ajustement

Sur le Tableau 2 apparaissent les résultats des ajustements obtenus avec le premier paramétrage (partie gauche du tableau), et le second (partie droite). Les paramètres L_-, L_+ et k du deuxième paramétrage ont été retransformés en les L_-, K et t_0 équivalents pour faciliter les comparaisons. Mathématiquement, on devrait parvenir aux mêmes résultats. Cela n'est pas strictement vrai. En dépit de la double précision utilisée par la bibliothèque NAG les itérations ne conduisent pas strictement à la même solution. Il a même été nécessaire avec le premier paramétrage d'introduire une borne supérieure à L_- , ce qui explique le chiffre rond obtenu,

égal à la borne (50.00). Il en va de même pour divers $P_{hj}^{(1)}$. Cela tient à l'accuité des problèmes de colinéarité. Tout un ensemble de valeurs différentes des paramètres conduit pratiquement à la même valeur minimale pour la fonction objectif.

7.3. Analyse de colinéarité globale/ paramétrage 1

Le Hessien apparaît dans le Tableau 3, son inverse standardisée dans le Tableau 4. Sous les hypothèses précédemment évoquées cet inverse standardisé correspond à la matrice de corrélation. On remarquera des très forts coefficients de corrélation, notamment entre K et $L_{..}$. Les chiffres du Tableau 4 sont arrondis, mais le chiffre exact obtenu est même supérieur en valeur absolue à 1. Ce résultat est absurde sur le plan théorique, mais tient à ce que les problèmes de colinéarité sont si forts que le Hessien frôle la singularité. Les problèmes numériques en deviennent considérables, notamment pour l'inversion. On ~~peut~~ en retrouvera plus loin des conséquences au niveau de l'ACP: H^{-1} , inverse du Hessien, devrait être une matrice diagonalisable, parceque symétrique positive. L'approximation obtenue ici ne l'est pas. Il apparaît aussi une valeur propre négative, quoique faible en valeur absolue (5% de la trace). Les configurations dans le plan des facteurs I et II feront de même apparaître certains points hors du cercle des corrélations, ce qui est théoriquement exclu, et relève donc de problèmes numériques.

On peut encore noter sur le Hessien que les dérivées secondes croisées, concernant des couples $P_{1,j}$ et $P_{2,j}$ sont nulles. Cela tient à la nature de la fonction objectif. Sur l'analyse globale il apparaît néanmoins des corrélations. Elles correspondent au fait que les paramètres $L_{..}$, K, t_o , a et b sont simultanément liés aux $P_{1,j}$ et $P_{2,j}$. Un examen attentif de la matrice de corrélation montre encore les valeurs négatives entre les P_{hj} d'une part, $P_{h(j-1)}$ et $P_{h(j+1)}$ d'autre part. Les erreurs d'estimation des P_{hj} créent surtout des transferts vers les classes adjacentes.

⁽¹⁾ De ce fait l'inverse du Hessien ne donne plus strictement la matrice de variance-covariance. Cela ne nous semble toutefois pas de nature à rendre impossible l'interprétation des colinéarités.

Tableau 3. Hessien Correspondant à 13 Variables
avec $\theta = (L, K, t_0, a, b, (p_{hj}))$, $h=1,2$; $j=1,2,3,4$

	L	K	t ₀	a	b	P ₁₁	P ₁₂	P ₁₃	P ₁₄	P ₂₁	P ₂₂	P ₂₃	P ₂₄
L	.29E+02												
K	.21E+04	.16E+06											
t ₀	-.34E+03	-.36E+05	.13E+05										
a	-.11E+02	-.18E+04	.44E+03	.25E+03									
b	.12E+01	-.49E+03	.17E+03	.13E+03	.76E+02								
P ₁₁	-.20E+03	-.14E+05	.15E+04	.18E+01	.16E+02	.14E+05							
P ₁₂	-.38E+03	-.30E+05	.40E+04	-.10E+03	-.18E+02	.66E+04	.10E+05						
P ₁₃	-.28E+03	-.18E+05	.15E+04	-.16E+02	.17E+02	.51E+04	.70E+04	.87E+04					
P ₁₄	-.13E+03	-.66E+04	.34E+03	-.13E+03	-.11E+03	.22E+04	.25E+04	.38E+04	.27E+04				
P ₂₁	-.31E+03	-.23E+05	.34E+04	-.20E+03	-.12E+03	.00E+00	.00E+00	.00E+00	.00E+00	.11E+05			
P ₂₂	-.20E+03	-.94E+04	-.64E+03	-.98E+02	-.12E+03	.00E+00	.00E+00	.00E+00	.00E+00	.73E+04	.98E+04		
P ₂₃	-.31E+03	-.19E+05	.14E+04	-.13E+03	-.13E+03	.00E+00	.00E+00	.00E+00	.00E+00	.48E+04	.59E+04	.69E+04	
P ₂₄	-.14E+03	-.78E+04	.51E+03	-.13E+03	-.13E+03	.00E+00	.00E+00	.00E+00	.00E+00	.25E+04	.28E+04	.37E+04	.26E+04

L'ACP fait apparaitre des pourcentages des premières valeurs propres par rapport à la trace, égaux respectivement à 48%, 20%, 12%, 7.5%, 6.8% et 3.9%. la concentration sur les premiers facteurs est donc très forte. Cela traduit une très intense colinéarité, qui explique au passage la lenteur des convergences dans la minimisation de la fonction objectif. D'un point de vue pratique on peut se demander s'il est raisonnable d'estimer plus de cinq, voire plus de trois, paramètres.

Tableau 4. Matrice de corrélation correspondant à 13 Variables
avec $\theta = (L_-, K, t_0, a, b, (p_{h,j}))$, $h=1,2$; $j=1,2,3,4$
écarts-types associés

	L_-	K	t_0	a	b	p_{11}	p_{12}	p_{13}	p_{14}	p_{21}	p_{22}	p_{23}	p_{24}
L_-	1.00												
K	-1.00	1.00											
t_0	-.92	.93	1.00										
a	.59	-.64	-.67	1.00									
b	-.66	.70	.68	-.95	1.00								
p_{11}	.02	.00	.02	.13	-.12	1.00							
p_{12}	-.07	-.02	-.11	-.61	.52	-.34	1.00						
p_{13}	.41	-.41	-.35	.55	-.64	.07	-.81	1.00					
p_{14}	.74	-.74	-.67	.16	-.17	-.09	.31	-.21	1.00				
p_{21}	.04	-.13	-.30	-.09	.02	.13	-.61	.05	-.01	1.00			
p_{22}	-.39	.47	.60	-.15	.26	-.14	.64	-.24	-.20	-.35	1.00		
p_{23}	.94	-.98	-.91	.46	-.61	.09	-.42	.45	.65	-.19	-.40	1.00	
p_{24}	-.43	.47	.47	-.32	.47	-.13	.57	-.39	-.14	.14	.21	-.65	1.00
σ	3.48	.04	.03	.19	.46	.01	.01	.03	.07	.01	.02	.07	.05

La figure 1 matérialise le problème. Le premier axe est surtout lié aux paramètres de croissance moyenne, en opposant L_- à t_0 et K. Les paramètres de variabilité individuelle de croissance, a et b, s'inscrivent assez bien dans le même schéma. b est opposé à L_- . Une variabilité individuelle accrue dans les âges élevés suppose un L_- moyen plus faible, à distributions de longueurs données. Cela est à rapprocher des travaux de POWELL (1979). L'opposition entre a et b tient à ce que pour un âge donné, on peut obtenir la même dispersion des longueurs en augmentant a si l'on diminue b, ou réciproquement.

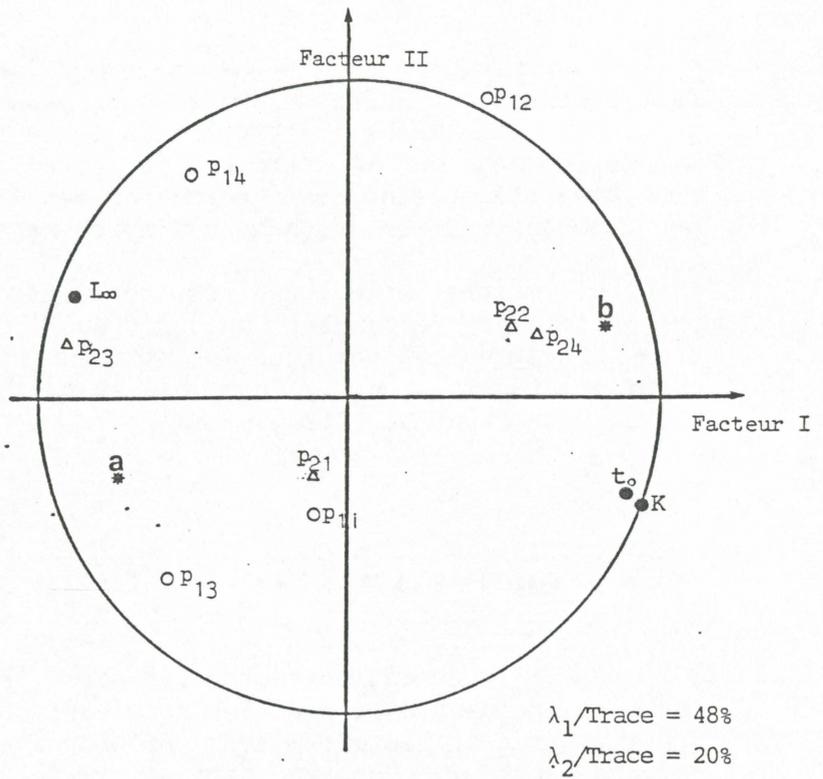


Figure 1 : Analyse en composantes principales de la matrice de corrélation globale.
Plan des facteurs 1 et 2

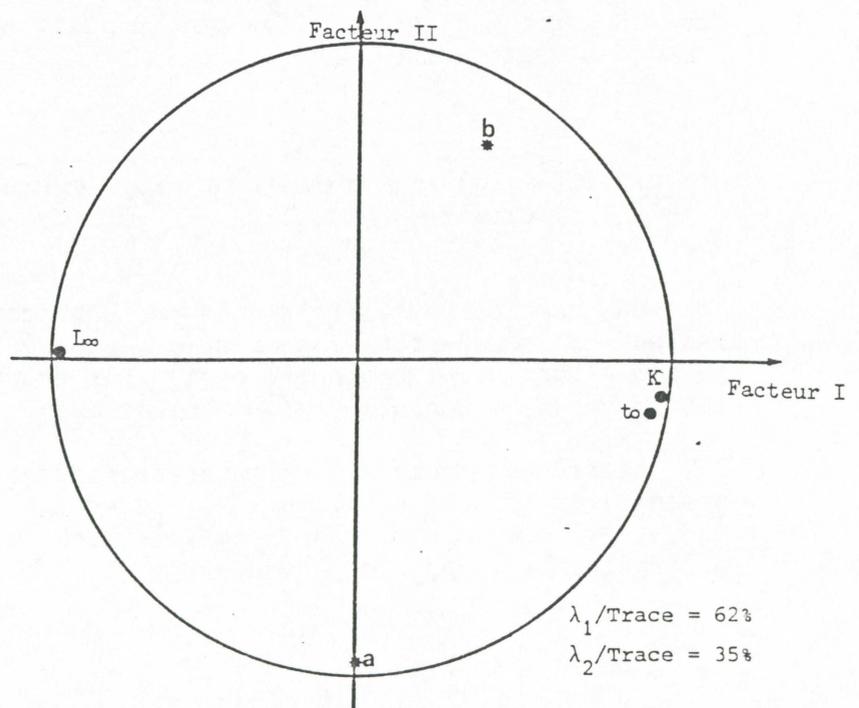


Figure 2 : Analyse en composantes principales limitée aux paramètres de croissance Facteurs I et II

On remarquera encore que si le deuxième axe (mais aussi en fait le troisième) affecte surtout les paramètres p_{nj} , ceux-ci ne sont pas globalement indifférents à l'axe 1. Comme le montre le simple examen de la matrice de corrélation, il existe des problèmes de colinéarité entre paramètres de croissance d'une part, paramètres démographiques d'autre part.

On notera enfin que les corrélations négatives entre proportions correspondant à des groupes d'âge adjacents sont transposées sur le graphique en positions associées opposées. Le cheminement de p_{12} en p_{13} , puis p_{13} et p_{14} montre ces oppositions, qui se retrouvent sur les p_{2j} .

7.4. Analyses partielles

Ces analyses ont porté sur les paramètres de croissance d'une part (en incluant a et b , liés à la variabilité individuelle), et d'autre part sur les paramètres démographiques. Dans les deux cas on considère le Hessien partiel. Cela équivaut en termes d'inference à supposer connus, et égaux à la valeur trouvée pour la solution retenue, les paramètres exclus de l'analyse. Cela se traduira entre autres par une réduction des variances, du moins en règle générale, et tout problèmes numérique mis à part. Le champ des incertitudes est en effet réduit puisque certains paramètres sont considérés comme connus.

7.4.1. Analyse réduite aux paramètres de croissance

La matrice de "corrélation" correspondante apparaît dans le Tableau 5. On retrouve de très fortes corrélations entre K , L_{-} et t_{0} d'une part, a et b d'autre part. Les écarts-types sont réduits (sauf pour t_{0}), conformément au principe énoncé plus haut.

La figure 2 fait apparaître une configuration particulièrement claire, où un axe "explique" K , L_{-} et t_{0} , tandis que le second rend compte de a . b étant à la fois négativement lié à L_{-} et a apparaît dans le premier quadrant.

Tableau 5. Matrice de corrélation
pour L_{∞} , K, t_0 , a et b
Ecart-types associés

	L_{∞}	K	t_0	a	b
L_{∞}	1.00				
K	-.99	1.00			
t_0	-.94	.96	1.00		
a	-.15	.19	.18	1.00	
b	-.35	.32	.29	-.80	1.00
σ	1.84	.03	.04	.17	.31

7.4.2. Analyse des paramètres démographiques

La matrice de "corrélation" obtenue apparait dans le tableau 6. Cette fois les p_{1j} et p_{2j} ne sont pas corrélés. Mathématiquement la structure du Hessien partiel l'implique, et son inverse correspond simplement aux deux matrices (4,4), inverses des Hessiens partiels associés aux p_{1j} d'une part, aux p_{2j} d'autre part. D'un point de vue statistique les corrélations ont disparu, car elles étaient dues au truchement des paramètres de croissance.

Après diagonalisation on retrouve très fidèlement des axes liés strictement à l'un ou l'autre des deux histogrammes. Cela transparait sur la Figure 3, où l'opposition entre les p_{n_j} et $p_{n_{j+1}}$ apparaît en outre nettement.

Un regard jeté aux écart-types montre leur accroissement avec l'âge. Cela légitime pour partie le souci d'introduire des contraintes supplémentaires dans les âges élevés.

Tableau 6. Matrice de Correlation pour
 $\{P_{hj}\}$, $h=1,2$; $j=1,2,3,4$
 Ecart-types associés

	P ₁₁	P ₁₂	P ₁₃	P ₁₄	P ₂₁	P ₂₂	P ₂₃	P ₂₄
P ₁₁	1.00							
P ₁₂	-.36	1.00						
P ₁₃	.01	-.63	1.00					
P ₁₄	-.10	.29	-.76	1.00				
P ₂₁	.00	.00	.00	.00	1.00			
P ₂₂	.00	.00	.00	.00	-.54	1.00		
P ₂₃	.00	.00	.00	.00	.04	-.52	1.00	
P ₂₄	.00	.00	.00	.00	-.11	.27	-.83	1.00
σ	.01	.02	.02	.03	.01	.02	.03	.04

7.5. Deuxième paramétrage - Analyse globale

La matrice de corrélation obtenue apparait dans le Tableau 7. Le plan des facteurs I et II constitue la Figure 4. Les changements majeurs ont évidemment trait aux paramètres de croissance. Ne pas retrouver des corrélations aussi fortes est a priori satisfaisant, mais le reparamétrage ne peut avoir accru le nombre de paramètres réellement estimables. Un problème numérique analogue à celui rencontré en 7.3 apparait d'ailleurs, car une valeur propre négative existe, égale à -6% de la trace. Les cinq premières valeurs propres sont respectivement égales à 31%, 25%, 14%, 10.4% et 7.8% de la trace. Le cumul de ces pourcentages mène à une concentration analogue à celle enregistrée précédemment. On remarquera encore sur la Figure 4 que les positions relatives des paramètres inchangés (a, b et les $\{P_{hj}\}$) sont très proches de celles apparaissant sur la Figure 1.

8. Conclusion

Le traitement des histogrammes de longueur par des méthodes mathématiques rigoureuses parait la seule voie d'avenir. Il ne

Tableau 7. Matrice de corrélation correspondant à 13 Variables
avec le paramétrage $\theta = (L_-, L_+, k, A, B, (p_{hj}))$, $h=1,2$; $j=1,2,3,4$
écarts-types associés

	L ₋	L ₊	k	a	b	P ₁₁	P ₁₂	P ₁₃	P ₁₄	P ₂₁	P ₂₂	P ₂₃	P ₂₄
L ₋	1.00												
L ₊	.30	1.00											
k	.54	1.00	1.00										
a	.20	.12	.42	1.00									
b	-.16	-.17	-.35	-.95	1.00								
P ₁₁	-.02	-.06	-.13	-.08	-.06	1.00							
P ₁₂	.16	.31	.73	-.42	.25	-.32	1.00						
P ₁₃	-.08	-.08	-.13	.34	-.37	-.04	-.72	1.00					
P ₁₄	.23	.71	.70	-.15	.20	-.12	.59	-.62	1.00				
P ₂₁	.28	-.16	.25	-.06	-.03	.12	-.64	.11	-.06	1.00			
P ₂₂	-.36	-.09	-.50	-.01	.13	-.10	.55	-.11	-.09	-.33	1.00		
P ₂₃	.23	.69	.93	.08	-.23	.03	.13	.05	.47	-.34	-.30	1.00	
P ₂₄	-.07	-.15	-.34	-.26	.41	-.06	.31	-.20	.03	.08	.12	-.68	1.00
σ	.16	.58	.01	.16	.37	.01	.01	.02	.05	.01	.01	.04	.04

faut cependant pas s'illusionner sur la quantité d'information que l'on peut extraire des seules distributions de longueur. En pratique le nombre de paramètres que l'on peut estimer est sévèrement limité par des problèmes de colinéarité. Pour analyser ces problèmes à l'échelle d'un couple de paramètres, reporter sur un graphique bidimensionnel les valeurs de la fonction objectif (SHEPHERD, 1985), et contourner les isoplèthes constitue une procédure irremplaçable. Dans un cadre plus complexe l'ACP de l'inverse du Hessien paraît constituer un outil d'analyse précieux.

Pour limiter les difficultés, l'incorporation de contraintes et/ou la réduction du nombre de paramètres constitue une première approche. Il est cependant vraisemblable que l'incorporation de données d'information externes, issues notamment de lectures d'âge sera souvent nécessaire (KIMURA et CHIKUNI, 1987).

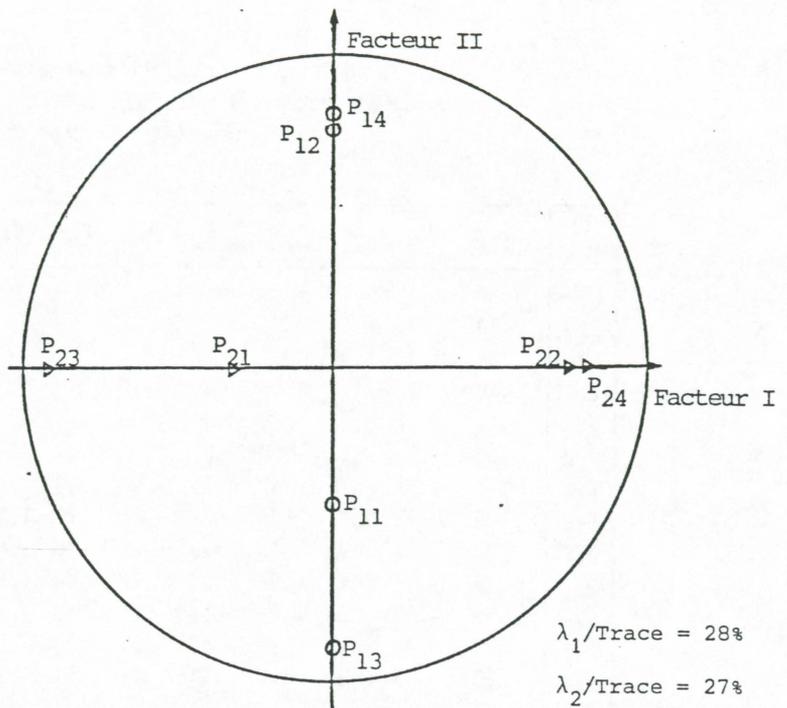


Figure 3: Analyse en composantes principales. Paramétrage 1
Colinéarité des structures démographiques.
Facteurs I et II

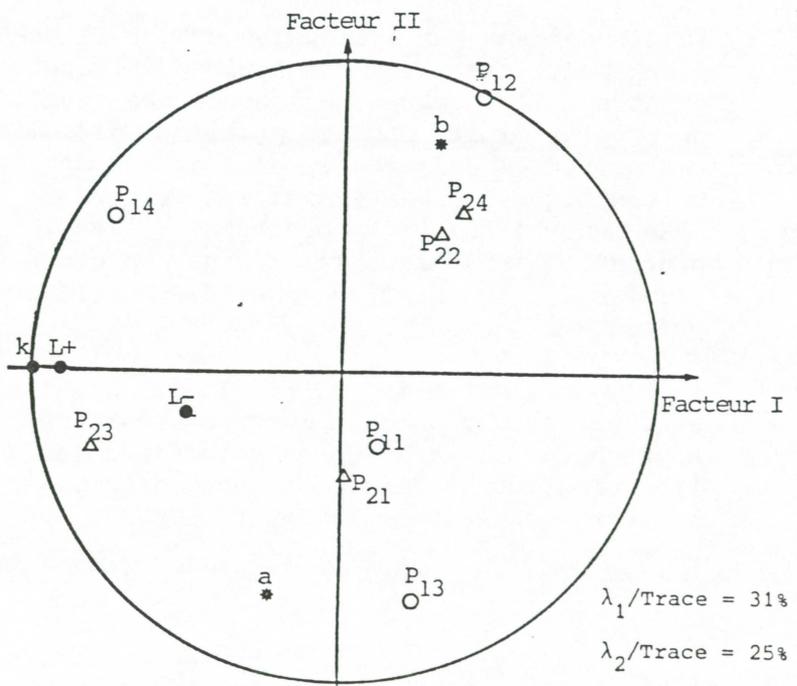


Figure 4 : Analyse en composantes principales. Paramétrage 2
Structure globale des colinéarités
Facteurs I et II

9. Bibliographie

- BREEN, P.A. and D.A. FOURNIER. 1984. A User's guide to estimating total Mortality rates from length frequency data with the method of FOURNIER and BREEN. Can. Tech. Rep. Fish. Aquat. Sci., 1239: 63p
- BRILLINGER, D.R. 1975. Time series, Data analysis and theory. Holt, Rinehart and Winston, Inc., New York. 500p
- FOURNIER, D.A. and P.A. BREEN. 1983. Estimation of abalone mortality rates with growth analysis. Trans. Am. Fish. Soc., 112: 403-411
- HASSELBLAD, V. 1966. Estimation of parameters for a mixture of normal distributions. Technometrics, 8(3): 431-444
- KIMURA, D.K. and S.CHIKUNI. 1987. Mixtures of empirical distributions: an iterative application of the age-length key. Biometrics, 43: 23-35
- PAULY, D. and G. GASCHUTS. 1979. A simple method for fitting oscillating growth data, with a program for pocket calculators. ICES C.M. 1979/G:24
- PAULY, D. 1985. A Review of the ELEFAN system for analysis of length frequency data in fish and aquatic invertebrates. In: PAULY, D. et G.R. MORGAN (Eds.)
- PAULY, D. and G.R. MORGAN (Eds.). 1987. Length based methods in fisheries research. Vol. 1. ICLARM Conference Proceedings 13. ICLARM, Manila, Philippines and Kuwait Institute for Scientific Research, Safat, Kuwait.
- POWELL, D.G. 1979. Estimation of mortality and growth parameters from the length frequency of a catch. Rapp. P.-v. Reun. Cons. Int. Explor. Mer, 175: 167-169
- SCHNUTE, J. and D.A. FOURNIER. 1980. A new approach to length frequency analysis: Growth structure. Can. J. Fish. Aquat. Sci., 37: 1337-1351
- SHEPHERD, J.G. 1985. A weakly parametric method for the analysis of length composition data. In: PAULY, D. and G.R. MORGAN (Eds.)

SPARRE, P. 1985. A method for estimation of growth, mortality and gear selection/recruitment parameters from multiple c.p.u.e. length frequency data. In: PAULY, D. and G.R. MORGAN (Eds.)